RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _10/750,891_

Source: _TEWO_

Date Processed by STIC: _4/15-05_

ENTERED

IFWO

```
RAW SEQUENCE LISTING              DATE: 04/15/2005
PATENT APPLICATION:  US/10/750,891    TIME: 10:37:54

Input Set : A:\Final sequence list-13173-00010-US.txt
Output Set: N:\CRF4\04152005\J750891.raw
```

```
 3 <110> APPLICANT: Puchta, Holger
 4        Biesgen, Christian
 6 <120> TITLE OF INVENTION: Recombination systems and methods for eliminating nucleic
acid
 7        sequences from genome of eukaryotic organisms
 9 <130> FILE REFERENCE: 53262-20031.00/13173-00010-US
11 <140> CURRENT APPLICATION NUMBER: US 10/750,891
12 <141> CURRENT FILING DATE: 2004-01-05
14 <150> PRIOR APPLICATION NUMBER: PCT/EP02/07281
15 <151> PRIOR FILING DATE: 2002-07-02
17 <150> PRIOR APPLICATION NUMBER: DE 101 31 786.7
18 <151> PRIOR FILING DATE: 2001-07-04
20 <160> NUMBER OF SEQ ID NOS: 30
22 <170> SOFTWARE: PatentIn version 3.3
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 788
27 <212> TYPE: DNA
28 <213> ORGANISM: Saccharomyces cerevisiae
30 <220> FEATURE:
31 <221> NAME/KEY: CDS
32 <222> LOCATION: (62)..(766)
33 <223> OTHER INFORMATION: open reading frame coding for I-SceI
35 <400> SEQUENCE: 1
36 ggatccagta ctgtacctag aatacaaaga agaggaagaa gaaacctcta cagaagaagt 60
37 g atg aaa aac atc aaa aaa aac cag gta atg aac ctg ggt ccg aac tct 109
38   Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro Asn Ser
39    1            5                 10                 15
40 aaa ctg ctg aaa gaa tac aaa tcc cag ctg atc gaa ctg aac atc gaa   157
41 Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn Ile Glu
42             20                 25                 30
43 cag ttc gaa gca ggt atc ggt ctg atc ctg ggt gat gct tac atc cgt   205
44 Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr Ile Arg
45             35                 40                 45
46 tct cgt gat gaa ggt aaa acc tac tgt atg cag ttc gag tgg aaa aac   253
47 Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp Lys Asn
48       50                 55                 60
49 aaa gca tac atg gac cac gta tgt ctg ctg tac gat cag tgg gta ctg   301
50 Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp Val Leu
51   65                 70                 75                 80
52 tcc ccg ccg cac aaa aaa gaa cgt gtt aac cac ctg ggt aac ctg gta   349
53 Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn Leu Val
54             85                 90                 95
55 atc acc tgg ggc gcc cag act ttc aaa cac caa gct ttc aac aaa ctg   397
56 Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn Lys Leu
```

RAW SEQUENCE LISTING   DATE: 04/15/2005
PATENT APPLICATION: US/10/750,891   TIME: 10:37:54

Input Set : A:\Final sequence list-13173-00010-US.txt
Output Set: N:\CRF4\04152005\J750891.raw

```
57            100              105              110
58 gct agc ctg ttc atc gtt aac aac aaa aaa acc atc ccg aac aac ctg  445
59 Ala Ser Leu Phe Ile Val Asn Asn Lys Lys Thr Ile Pro Asn Asn Leu
60         115              120              125
61 gtt gaa aac tac ctg acc ccg atg tct ctg gca tac tgg ttc atg gat  493
62 Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe Met Asp
63     130              135              140
64 gat ggt ggt aaa tgg gat tac aac aaa aac tct acc aac aaa tcg atc  541
65 Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Thr Asn Lys Ser Ile
66 145              150              155              160
67 gta ctg aac acc cag tct ttc act ttc gaa gaa gta gaa tac ctg gtt  589
68 Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Glu Tyr Leu Val
69              165              170              175
70 aag ggt ctg cgt aac aaa ttc caa ctg aac tgt tac cta aaa atc aac  637
71 Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Leu Lys Ile Asn
72         180              185              190
73 aaa aac aaa ccg atc atc tac atc gat tct atg tct tac ctg atc ttc  685
74 Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu Ile Phe
75         195              200              205
76 tac aac ctg atc aaa ccg tac ctg atc ccg cag atg atg tac aaa ctg  733
77 Tyr Asn Leu Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr Lys Leu
78     210              215              220
79 ccg aac act atc tcc tcc gaa act ttc ctg aaa taataagtcg agtactggat  786
80 Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys
81 225              230              235
82 cc                                                            788
84 <210> SEQ ID NO: 2
85 <211> LENGTH: 235
86 <212> TYPE: PRT
87 <213> ORGANISM: Saccharomyces cerevisiae
89 <400> SEQUENCE: 2
90 Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro Asn Ser
91  1               5                  10              15
92 Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn Ile Glu
93         20               25               30
94 Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr Ile Arg
95         35               40               45
96 Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp Lys Asn
97     50               55               60
98 Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp Val Leu
99 65               70               75               80
100 Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn Leu Val
101             85               90               95
102 Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn Lys Leu
103         100              105              110
104 Ala Ser Leu Phe Ile Val Asn Asn Lys Lys Thr Ile Pro Asn Asn Leu
105         115              120              125
106 Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe Met Asp
107     130              135              140
```

```
RAW SEQUENCE LISTING                    DATE: 04/15/2005
PATENT APPLICATION:   US/10/750,891     TIME: 10:37:54

Input Set : A:\Final sequence list-13173-00010-US.txt
Output Set: N:\CRF4\04152005\J750891.raw

108 Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Thr Asn Lys Ser Ile
109 145                 150                 155                 160
110 Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Glu Tyr Leu Val
111                 165                 170                 175
112 Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Leu Lys Ile Asn
113                 180                 185                 190
114 Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu Ile Phe
115         195                 200                 205
116 Tyr Asn Leu Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr Lys Leu
117     210                 215                 220
118 Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys
119 225                 230                 235
121 <210> SEQ ID NO: 3
122 <211> LENGTH: 746
123 <212> TYPE: DNA
124 <213> ORGANISM: Chlamydomonas applanata
126 <220> FEATURE:
127 <221> NAME/KEY: CDS
128 <222> LOCATION: (54)..(737)
129 <223> OTHER INFORMATION: open reading frame of I-ChuI with nuclear location
130       signal
132 <220> FEATURE:
133 <221> NAME/KEY: misc_feature
134 <222> LOCATION: (54)..(83)
135 <223> OTHER INFORMATION: coding for nuclear location signal
137 <400> SEQUENCE: 3
138 ctcgagtacc tagaatacaa agaagaggaa gaagaaactc tatagaagaa gcc atg    56
139                                                           Met
140                                                           1
141 ggt cca aag aaa aag aga aag gtt atc atg tca tta aca caa caa caa   104
142 Gly Pro Lys Lys Lys Arg Lys Val Ile Met Ser Leu Thr Gln Gln Gln
143             5                   10                  15
144 aaa gac tta att ttc gga tct cta ctg ggt gat gga aat tta caa act   152
145 Lys Asp Leu Ile Phe Gly Ser Leu Leu Gly Asp Gly Asn Leu Gln Thr
146         20                  25                  30
147 ggt tca gta ggt agg act tgg cgc tat cga gcg ctc cat aaa agt gag   200
148 Gly Ser Val Gly Arg Thr Trp Arg Tyr Arg Ala Leu His Lys Ser Glu
149     35                  40                  45
150 cat cag aca tac tta ttt cat aag tat gaa atc tta aag ccg ctt tgt   248
151 His Gln Thr Tyr Leu Phe His Lys Tyr Glu Ile Leu Lys Pro Leu Cys
152 50                  55                  60                  65
153 ggc gaa aat act ctc cca aca gaa agt ata gtg ttc gac gaa aga aca   296
154 Gly Glu Asn Thr Leu Pro Thr Glu Ser Ile Val Phe Asp Glu Arg Thr
155                 70                  75                  80
156 aac aag gag gtt aaa cgt tgg ttt ttc aac aca tta acc aat cct tcc   344
157 Asn Lys Glu Val Lys Arg Trp Phe Phe Asn Thr Leu Thr Asn Pro Ser
158                 85                  90                  95
159 tta aaa ttc ttc gca gac atg ttc tac aca tat gac caa aac aca caa   392
160 Leu Lys Phe Phe Ala Asp Met Phe Tyr Thr Tyr Asp Gln Asn Thr Gln
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/750,891

DATE: 04/15/2005
TIME: 10:37:54

Input Set : A:\Final sequence list-13173-00010-US.txt
Output Set: N:\CRF4\04152005\J750891.raw

```
161         100                    105                    110
162 aaa tgg gtt aaa gat gta cct gta aag gtt caa aca ttc tta act cct    440
163 Lys Trp Val Lys Asp Val Pro Val Lys Val Gln Thr Phe Leu Thr Pro
164     115                    120                    125
165 caa gct tta gca tac ttt tat ata gac gat gga gcg tta aaa tgg ctt    488
166 Gln Ala Leu Ala Tyr Phe Tyr Ile Asp Asp Gly Ala Leu Lys Trp Leu
167 130                    135                    140                    145
168 aat aag tct aac gct atg caa att tgt act gaa agt ttc agt caa ggg    536
169 Asn Lys Ser Asn Ala Met Gln Ile Cys Thr Glu Ser Phe Ser Gln Gly
170                    150                    155                    160
171 ggc acg att cgg atc caa aaa gca cta aaa acg ctc tat aat att gat    584
172 Gly Thr Ile Arg Ile Gln Lys Ala Leu Lys Thr Leu Tyr Asn Ile Asp
173                    165                    170                    175
174 aca acg ttg aca aaa aaa act cta caa gac ggc aga att ggc tat cgt    632
175 Thr Thr Leu Thr Lys Lys Thr Leu Gln Asp Gly Arg Ile Gly Tyr Arg
176         180                    185                    190
177 ata gct att cct gaa gcc agt agc ggt gct ttt cgt gaa gtc att aaa    680
178 Ile Ala Ile Pro Glu Ala Ser Ser Gly Ala Phe Arg Glu Val Ile Lys
179     195                    200                    205
180 cct ttt cta gtt gat tgt atg aga tac aaa gtt tct gat ggc aat aaa    728
181 Pro Phe Leu Val Asp Cys Met Arg Tyr Lys Val Ser Asp Gly Asn Lys
182 210                    215                    220                    225
183 ggc cac ctt tagctcgag                                                746
184 Gly His Leu
```

187 <210> SEQ ID NO: 4
188 <211> LENGTH: 228
189 <212> TYPE: PRT
190 <213> ORGANISM: Chlamydomonas applanata

192 <400> SEQUENCE: 4

```
193 Met Gly Pro Lys Lys Lys Arg Lys Val Ile Met Ser Leu Thr Gln Gln
194 1               5                    10                    15
195 Gln Lys Asp Leu Ile Phe Gly Ser Leu Leu Gly Asp Gly Asn Leu Gln
196                    20                    25                    30
197 Thr Gly Ser Val Gly Arg Thr Trp Arg Tyr Arg Ala Leu His Lys Ser
198         35                    40                    45
199 Glu His Gln Thr Tyr Leu Phe His Lys Tyr Glu Ile Leu Lys Pro Leu
200     50                    55                    60
201 Cys Gly Glu Asn Thr Leu Pro Thr Glu Ser Ile Val Phe Asp Glu Arg
202 65                    70                    75                    80
203 Thr Asn Lys Glu Val Lys Arg Trp Phe Phe Asn Thr Leu Thr Asn Pro
204                    85                    90                    95
205 Ser Leu Lys Phe Phe Ala Asp Met Phe Tyr Thr Tyr Asp Gln Asn Thr
206         100                    105                    110
207 Gln Lys Trp Val Lys Asp Val Pro Val Lys Val Gln Thr Phe Leu Thr
208         115                    120                    125
209 Pro Gln Ala Leu Ala Tyr Phe Tyr Ile Asp Asp Gly Ala Leu Lys Trp
210         130                    135                    140
211 Leu Asn Lys Ser Asn Ala Met Gln Ile Cys Thr Glu Ser Phe Ser Gln
212 145                    150                    155                    160
```

```
                    RAW SEQUENCE LISTING              DATE: 04/15/2005
                    PATENT APPLICATION:  US/10/750,891   TIME: 10:37:54

                    Input Set : A:\Final sequence list-13173-00010-US.txt
                    Output Set: N:\CRF4\04152005\J750891.raw
```

```
213 Gly Gly Thr Ile Arg Ile Gln Lys Ala Leu Lys Thr Leu Tyr Asn Ile
214             165             170             175
215 Asp Thr Thr Leu Thr Lys Lys Thr Leu Gln Asp Gly Arg Ile Gly Tyr
216             180             185             190
217 Arg Ile Ala Ile Pro Glu Ala Ser Ser Gly Ala Phe Arg Glu Val Ile
218         195             200             205
219 Lys Pro Phe Leu Val Asp Cys Met Arg Tyr Lys Val Ser Asp Gly Asn
220     210             215             220
221 Lys Gly His Leu
222 225
225 <210> SEQ ID NO: 5
226 <211> LENGTH: 582
227 <212> TYPE: DNA
228 <213> ORGANISM: Chlamydomonas reinhardtii
230 <220> FEATURE:
231 <221> NAME/KEY: CDS
232 <222> LOCATION: (55)..(573)
233 <223> OTHER INFORMATION: openreading frame coding for I-CreI with nuclear
234       location signal
236 <220> FEATURE:
237 <221> NAME/KEY: misc_feature
238 <222> LOCATION: (55)..(84)
239 <223> OTHER INFORMATION: coding for nuclear location signal
241 <400> SEQUENCE: 5
242 ctcgagtacc tagaatacaa agaagaggaa gagaaacctc taccagaaga agcc atg   57
243                                                               Met
244                                                                 1
245 ggt cca aag aaa aag aga aag gtt atc atg aat aca aaa tat aat aaa   105
246 Gly Pro Lys Lys Lys Arg Lys Val Ile Met Asn Thr Lys Tyr Asn Lys
247             5               10              15
248 gag ttc tta ctc tac tta gca ggg ttt gta gac ggt gac ggt agc ata   153
249 Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile
250         20              25              30
251 atc gct caa att aag cct aat cag tct tat aaa ttt aag cat cag cta   201
252 Ile Ala Gln Ile Lys Pro Asn Gln Ser Tyr Lys Phe Lys His Gln Leu
253     35              40              45
254 tca ctc gcg ttc caa gtc acg caa aag aca cag aga cgt tgg ttt tta   249
255 Ser Leu Ala Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe Leu
256 50              55              60              65
257 gac aaa tta gtg gat gaa att ggg gtt ggt tat gta aga gat agg ggt   297
258 Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Arg Asp Arg Gly
259             70              75              80
260 agc gtt tcg gat tat att cta agc gaa atc aag cct ttg cat aat ttt   345
261 Ser Val Ser Asp Tyr Ile Leu Ser Glu Ile Lys Pro Leu His Asn Phe
262             85              90              95
263 tta aca caa cta caa cct ttt cta aaa cta aaa caa aaa caa gca aat   393
264 Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
265         100             105             110
266 tta gtt tta aaa att att gaa caa ctt ccg tca gca aaa gaa tcc ccg   441
```

VERIFICATION SUMMARY
PATENT APPLICATION: US/10/750,891

DATE: 04/15/2005
TIME: 10:37:55

Input Set : **A:\Final sequence list-13173-00010-US.txt**
Output Set: **N:\CRF4\04152005\J750891.raw**